

## Statistical Recalibration of GCM Forecasts over Southern Africa Using Model Output Statistics

WILLEM A. LANDMAN AND LISA GODDARD

*International Research Institute for Climate Prediction, Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York*

(Manuscript received 5 September 2001, in final form 18 January 2002)

### ABSTRACT

A technique for producing regional rainfall forecasts for southern Africa is developed that statistically maps or “recalibrates” large-scale circulation features produced by the ECHAM3.6 general circulation model (GCM) to observed regional rainfall for the December–February (DJF) season. The recalibration technique, model output statistics (MOS), relates archived records of GCM fields to observed DJF rainfall through a set of canonical correlation analysis (CCA) equations. After screening several potential predictor fields, the 850-hPa geopotential height field is selected as the single predictor field in the CCA equations that is subsequently used to produce MOS-recalibrated rainfall patterns. The recalibrated forecasts outscore area-averaged GCM-simulated rainfall anomalies, as well as forecasts produced using a simple linear forecast model. The MOS recalibration is applied to two sets of GCM experiments: for the “simulation” experiment, simultaneous observed sea surface temperature (SST) serves as the lower boundary forcing; for the “hindcast” experiment, the prescribed SSTs are obtained by persisting the previous month’s SST anomaly through the forecast period. Pattern analyses performed on the predictor–predictand pairs confirm a robust relationship between the GCM 850-hPa height fields and the rainfall fields. The structure and variability of the large-scale circulation is well characterized by the GCM in both simulation and hindcast mode. Measures of retroactive skill for a 9-yr independent period (1991/92–1999/2000) using the hindcast MOS are obtained for both deterministic and probabilistic forecasts, suggesting that a probabilistic representation of MOS forecasts is potentially more valuable. Finally, MOS is employed to investigate its potential to downscale the GCM large-scale circulation to more specific forecasts of land surface characteristics such as streamflow.

### 1. Introduction

Over the most recent decade major improvements in understanding the predictability of southern Africa’s seasonal rainfall has emerged (Cane et al. 1994; Hastenrath et al. 1995; Barnston and Smith 1996; Hunt 1997; Mason et al. 1996, 1999; Jury 1996; Mason 1998; Mattes and Mason 1998; Makarau and Jury 1997; Jury et al. 1999; Landman and Mason 1999b; Landman and Tennant 2000; Landman et al. 2001a). Of the two approaches currently used to determine the future behavior of the climate system; namely, a purely empirical-statistical approach and a dynamical approach using the first principles of the processes governing the climate system, the latter has received far less investigation as a seasonal forecasting technique for southern Africa. However, recently the emphasis of research for the region has begun to shift toward the use of more sophisticated forecast schemes involving the use of dynamical models based on first principles. These dynamical mod-

els of the climate system, or general circulation models (GCMs), are already being used extensively in seasonal forecasting globally (e.g., Palmer and Anderson 1994; Ji et al. 1994; Hunt 1997; Mason et al. 1999; Gates et al. 1999; Goddard et al. 2001).

Although GCMs, commonly configured with an effective resolution of 200–300 km, demonstrate skill at global or even continental scale, they are unable to represent local subgrid features. Over many parts of the world, including southern Africa, GCMs typically overestimate rainfall amounts and often spatially distort patterns of rainfall variability (Joubert and Hewitson 1997; Mason and Joubert 1997). Such systematic biases suggest the need to recalibrate, or downscale, GCM simulations (Karl et al. 1990; Crane and Hewitson 1998; Cui et al. 1995; Solman and Nuñez 1999; Zorita and von Storch 1999; Busuioc et al. 2001) to regional level over southern Africa (Landman and Tennant 2000; Landman et al. 2001a). Semiempirical relationships exist between observed large-scale circulation and rainfall, and assuming that these relationships are valid under future climate conditions and also that the large-scale structure and variability is well characterized by GCMs, mathematical equations can be constructed to predict

---

*Corresponding author address:* Dr. Willem A. Landman, South African Weather Service, Private Bag X097, Pretoria 0001, South Africa.  
E-mail: willem@weathersa.co.za

local precipitation from simulated large-scale circulation (Wilby and Wigley 2000).

The model output statistics (MOS) recalibration presented here focuses on the December–February (DJF) season for rainfall, which marks a particularly relevant season for southern Africa both meteorologically and societally. Most of southern Africa lies just outside of the Tropics where seasonal predictability is generally lower because of the greater inherent chaotic variability of the extratropical atmosphere (Palmer and Anderson 1994). However, during the peak of the austral summer rainfall season, DJF, tropical atmospheric circulation dominates. The dominance of tropical atmospheric circulation over the summer rainfall region means that local variations in atmospheric heating, such as from prescribed SSTs, exert a more direct influence on the atmospheric circulation, and the chaotic influences from atmospheric internal variability are at a minimum, leading to greater potential predictability during DJF. GCMs demonstrate skill in simulating circulation patterns over the Tropics (Hunt 1997; Shukla 1998; Stockdale et al. 1998; Mason et al. 1999), creating the prospect of useful GCM forecast skill simulating DJF circulation over southern Africa when a tropical atmosphere dominates (Mason et al. 1996). Of economic and social concern to southern Africa, this season is important for agriculture, especially for crop farmers, as tasselling and grain filling typically occurs during January and February (Mjelde et al. 1997).

In this paper, the ability of the proposed recalibration forecast system to improve upon the GCM-produced rainfall forecasts is assessed. Tests are performed to see if the scheme can improve also on the skill of a computationally simpler and less expensive model that relates rainfall to global SSTs. The robustness of the relationship between predictor and predictand fields is investigated over training periods of varying lengths to investigate the degree to which the statistical relationships of the recalibration equations are likely to be valid under future climate conditions. As a final step, the real-time operational forecast skill of a GCM as part of an operational forecast system is investigated over a 9-yr independent test period. With the skill and robustness of this approach established for regional rainfall forecasts, the use of the recalibration as an applications forecast tool is investigated for streamflow forecasts at the inlets of a number of dams in the region.

## 2. Data

### a. Sea surface temperatures

Reynolds's reconstructed sea surface temperature (SST) data (Smith et al. 1996) serve as the boundary forcing for the ECHAM3.6 GCM experiments and also for constructing and running the empirical-statistical forecast model. The SST data covers the period 1950 to the present. The data are interpolated from its original

$2^\circ \times 2^\circ$  resolution to T42 (approximately  $2.8^\circ$  horizontal resolution) before being applied to the GCM.

### b. Rainfall and streamflow

DJF rainfall totals for close to 600 southern African rainfall stations, including South Africa, Namibia, Lesotho, and Botswana, are obtained for the period 1950–2000. Regional rainfall indices are computed for nine homogeneous rainfall regions as shown in Fig. 1 (Mason 1998; Landman et al. 2001a). The austral summer rainfall regions are Transkei; the KwaZulu-Natal coast; Lowveld; the northeastern, central, and western interior regions; and northern Namibia/western Botswana. The southwestern Cape is predominantly an austral winter rainfall region, and the south coast gets rain throughout the year; therefore, these two regions will not be included in the analyses presented in this paper, as these regions are not strongly influenced by tropical circulation in DJF and as a result exhibit lower predictability.

Naturalized streamflow, derived by eliminating the influence of developments within a catchment from the observed streamflow of that catchment, are available from 1920 to the first few months of 1995. The streamflow, which are 3-month mean DJF values and measured in cubic meters per second, are considered for six dams of the Vaal and upper Tugela River systems (Landman et al. 2001b). These systems are located over the southeast of the northeastern interior shown in Fig. 1.

## 3. Methods

### a. The GCM

Two sets of integrations were performed using the ECHAM3.6 GCM (Deutsches Klimarechenzentrum 1992). The first ensemble of 10 runs was forced with simultaneous observed SSTs for the DJF season 1950/51–1999/2000. The resulting GCM fields are referred to as *simulation mode* fields. At initialization, ensemble members differ from each other by one model day at the beginning of the integration. The second ensemble contains five runs and was forced with November SST anomalies persisted on top of the monthly varying annual cycle of SST for the DJF season of 1970/71–1999/2000. The GCM fields from this ensemble are referred to as *hindcast mode*. The initial conditions for each of the hindcast ensemble members at the beginning of December in each year were taken from restart files of the simulations. Thus, the two sets of integrations have identical initial conditions and differ only in their prescribed SST anomalies. Figure 2 compares the observations with the 10-member ensemble mean simulation rainfall anomalies for the period 1970/71–1999/2000, as this is the period common to both sets of integrations. The data have been averaged over the grid points located within each of the regions specified in Fig. 1. This comparison, represented by the correlation value printed in

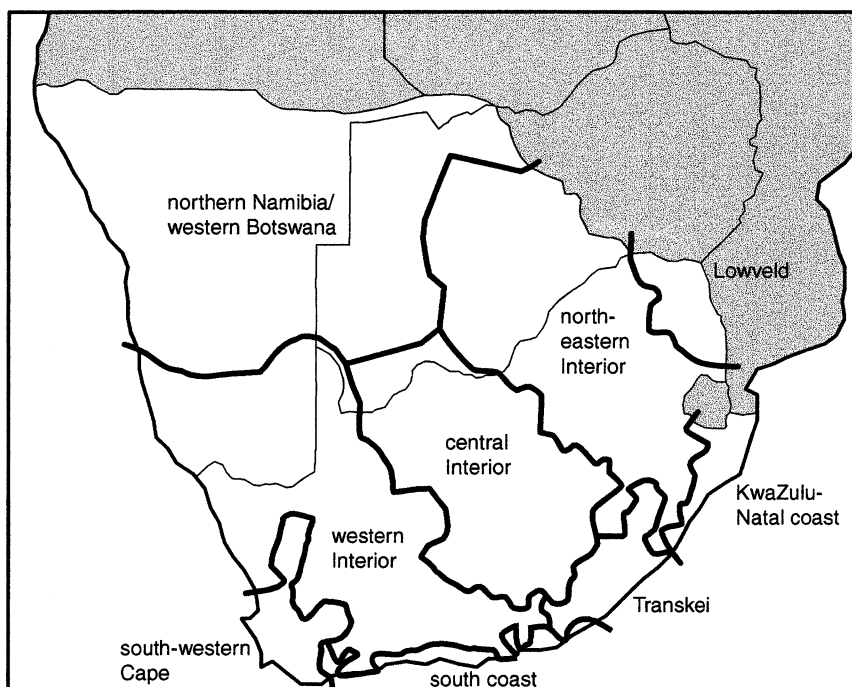


FIG. 1. The rainfall regions used in the study. Countries shaded gray are not included.

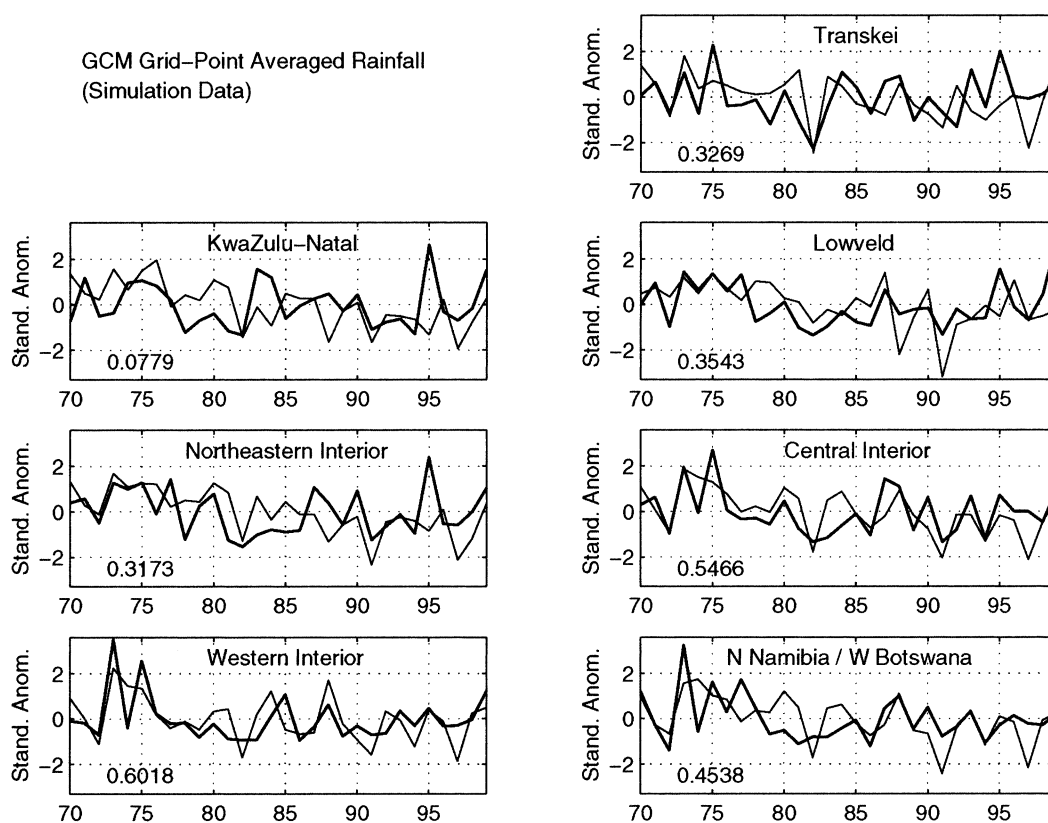


FIG. 2. Simulation ensemble mean DJF rainfall standardized anomalies (thin line), obtained from first averaging over the grid points located within each of the regions specified in Fig. 1, vs the observed DJF rainfall standardized anomalies (thick line) for each of the seven summer rainfall regions. The correlation between the two time series of each region is in the bottom left-hand corner.

the lower left-hand corner of each plot, forms the baseline skill level that must be outscored by the recalibration method that uses simulation mode fields, in order to justify its use.

#### *b. The recalibration method*

Statistical correction of GCM output is often necessary because significant biases occur between the real world and its modeled representation. The typical resolution of dynamical models used for seasonal climate prediction (approximately 200 km) often leads to poor simulation of small-scale effects that are important to local climate, such as the interaction between atmospheric circulation and detailed topography. Variables such as rainfall may therefore not be accurately represented by these models. However, statistical approaches can be utilized to construct relationships between the desired forecast quality and variables, such as large-scale circulation and moisture, that are simulated more accurately. Furthermore, quantities such as streamflow that are not simulated by GCMs, can be simulated through statistical interpretation.

The statistical approach used here to develop equations relating the GCM quantities to a forecast quantity is called model output statistics (MOS; Wilks 1995). This approach is normally preferred above perfect prognosis recalibration schemes (i.e., Landman et al. 2001a,b) because it can compensate for systematic errors in the GCM fields directly in the regression equations. These errors can be overcome because MOS uses predictor values from the GCM in both the development and forecast stages. Therefore, MOS forecast equations require a developmental dataset that consists of historical records of the predictand (i.e., rainfall) as well as archived records of the GCM predictor fields (i.e., large-scale circulation) for the same season. The time lag in MOS forecasts is therefore incorporated in the GCM forecasts. Separate MOS forecast equations must be developed for different forecast projections, owing to the decrease in skill of GCM forecast fields with increasing lead time.

Canonical correlation analysis (CCA) is the mathematical technique used in the paper to set up the MOS recalibration equations. This technique has been used before to simulate rainfall and streamflow over southern Africa from GCM output (Landman and Tennant 2000; Landman et al. 2001a,b). The first step in designing the CCA regression equations is to design the optimal MOS model. Empirical orthogonal function (EOF) analysis is performed on the predictor and predictand sets (Barnston 1994), and the number of modes to be retained in the CCA eigenanalysis problem is determined using cross-validated (Michaelsen 1987; Elsner and Schmertmann 1994) skill sensitivity tests. For cross validation, the value that is to be predicted is omitted from the training period. Here, only 1 yr is removed from the training period. The 1-yr-out cross-validation design is

adequate here because all of the regions have little or no autocorrelation at a lag of 1 yr: all the correlation values are below 0.2 with 6 of the 7 below 0.0, indicating that knowledge of the climate for year  $x - 1$  and year  $x + 1$  does not give information about the climate of year  $x$ .

The GCM large-scale patterns considered for the MOS model are the DJF GCM simulation fields of geopotential heights (850, 700, 500, and 200 hPa), thickness (850–500 and 500–200 hPa) and moisture (700-hPa specific humidity). To select the best predictor(s) from the array of GCM fields, the screening process called *forward selection* (Wilks 1995) is employed. The number of retained predictor and predictand EOF modes of the fields or combination of fields that produced the highest average cross-validated correlation for the austral summer rainfall regions of southern Africa is subsequently identified. The number of CCA modes is determined by using the Guttman–Kaiser criterion (Jackson 1991), but with a minimum of two CCA modes. The optimal equations are constructed for a cross-validated period of 29 yr from DJF 1970/71 to DJF 1998/99.

#### *c. Estimating MOS operational forecast skill*

Performance of the CCA recalibration system is first estimated through cross-validated correlation skill. Although cross-validated correlation values may be high and significant, some of that skill may be artificial due to biases in the validation year(s) relative to the training period (Barnston and van den Dool 1993), such as those due to trends. Model skill should therefore be determined over a test period that is independent of the training period, and should involve evaluation of predictions compared to observations excluding any information following the target year. Such a forecast validation system is referred to as retroactive, and is constructed here by considering forecasts for DJF during the 9-yr period of 1991/92–1999/2000. Details of a retroactive analysis can be found in Landman et al. (2001a). In estimating the skill of the DJF rainfall MOS forecasts each of the predicted and observed fields is separated into its own three equiprobable groups defining above-normal, near-normal, or below-normal conditions. These forecasts can be treated deterministically based on which category is assigned the highest probability or probabilistically considering the probabilities of each category. The categorized forecasts are compared with that of the observed in order to calculate the skill of the forecast system. The scores considered here are linear error in probability space (LEPS) scores (Ward and Folland 1991) for the retroactive deterministic forecasts, and the ranked probability skill score (Wilks 1995) for retroactive probabilistic forecasts.

#### *d. Significance estimation of skill levels*

The significance of the skill measures is determined through a Monte Carlo process (Livezey and Chen



1983). For cross-validation correlation significance, “predicted” and “observed” time series for varying number of training periods are randomly created by re-sampling the real data. For each of the training periods, a sequence of 1000 time series are randomly created and correlated. The 900th, 950th, and 990th ranked correlation values correspond to the 90%, 95%, and 99% levels of confidence, respectively, for the particular training period. This process is performed 100 times, then averaged at each of the three confidence levels and for a varying number of training periods. A one-tailed test is performed because only positive cross-validation correlation values are considered to be of significance. The 95% level of confidence is selected as the preferred level of significance. For example, the 21-yr training period is associated with a 95% confidence level threshold correlation of 0.37, the 24-yr period with 0.34, the 27-yr period with 0.32, and the 30-yr period with 0.31. A Monte Carlo process is similarly followed to determine the significance levels of correlation differences.

In order to determine the level of statistical significance of the LEPS scores, a Monte Carlo test is again performed by randomly creating above-normal, near-normal, and below-normal predicted and observed rainfall categories using a normally distributed random number generator. The LEPS scores are subsequently calculated and the procedure is repeated 1000 times. This process is done 500 times in order to calculate an average LEPS score associated with the respective significance levels. For a 9-yr period, the 90% confidence level corresponds to a LEPS score of 37%; the 95% level with a LEPS of 48%; the 99% level with a LEPS of 66%.

#### *e. CCA diagnostics and significance of spatial patterns*

The diagnostic features of CCA (spatial patterns and time scores) are discussed in detail in Barnett and Preisendorfer (1987), and are used here to give an indication of the physical processes that are responsible for the skill of the MOS recalibration. The spatial patterns or maps show the association between the predictor and the predictand and their respective canonical coefficients or time scores. These time scores and spatial patterns together indicate the association between the GCM large-scale circulation and the regional rainfall. The combination of significant CCA modes ultimately leads to the predictor and predictand relationship captured in the MOS equations. The statistical significance of CCA predictor map correlations is also obtained through a Monte Carlo process by randomly resampling the predictor canonical time scores and correlating the resampled time scores with each of the gridpoint values of the original GCM-derived large-scale field.

## **4. Results**

### *a. The best predictor large-scale field*

Of all the candidate, or combination of candidate GCM simulation mode fields considered, the 850-hPa geopotential height field produces the highest average cross-validation correlation over the austral summer rainfall regions. In descending order of average correlation the next five best candidate predictor fields are 1) the combined 850-hPa geopotential heights and GCM rainfall; 2) the combined 850- and 700-hPa geopotential heights; 3) the combined 850- and 700-hPa geopotential heights and the GCM rainfall; 4) the 700-hPa geopotential heights; and 5) the combined GCM rainfall and 700-hPa moisture. For the selected field of GCM 850-hPa geopotential heights, two predictand (69% of the variance) and two predictor (84% of the variance) EOF modes are used in the CCA-based MOS model, resulting in the selection of two dominant CCA modes. The differences in the average cross-validation correlations associated with each of the GCM fields, or combination of fields, are very small in most cases. By comparing the time scores of the first two EOF modes of the 850-hPa geopotential height field with those of the other candidate fields or combination of fields that produced the next best average correlation values, significant agreement is found (correlations > 0.9). This agreement indicates that there would be very little difference in skill of the MOS forecasts using the other fields or combination of fields as predictors instead of only the 850-hPa geopotential heights. Therefore, the inclusion of these latter fields is redundant in setting up the optimal MOS equations.

### *b. Performance of the MOS model using simulation data*

In general, the MOS model performed well over southern Africa; all of the cross-validation correlations of the simulation-MOS forecasts of the summer rainfall regions are higher than 0.31, which is the threshold for the 95% level of confidence (Fig. 3). This good performance is due in part to the ability of the GCM to reproduce the observed teleconnection between El Niño–La Niña events and southern African rainfall, which typically leads to drier conditions during El Niño events and wetter conditions during La Niña events. Thus, skillful simulations are found for most of the austral summer rainfall regions, especially during the excessively wet conditions associated with the La Niña years of 1973/74 and 1975/76 and for the dry conditions of the 1982/83 El Niño and of the early 1990s El Niño events. However, the atmospheric response of the GCM tends to overreact to El Niño–La Niña events in the tropical Pacific, which may be partly responsible for the seemingly linear El Niño–La Niña signal in the MOS model. For example, the drought forecast for the very strong El Niño event of 1982/83 is overestimated, and

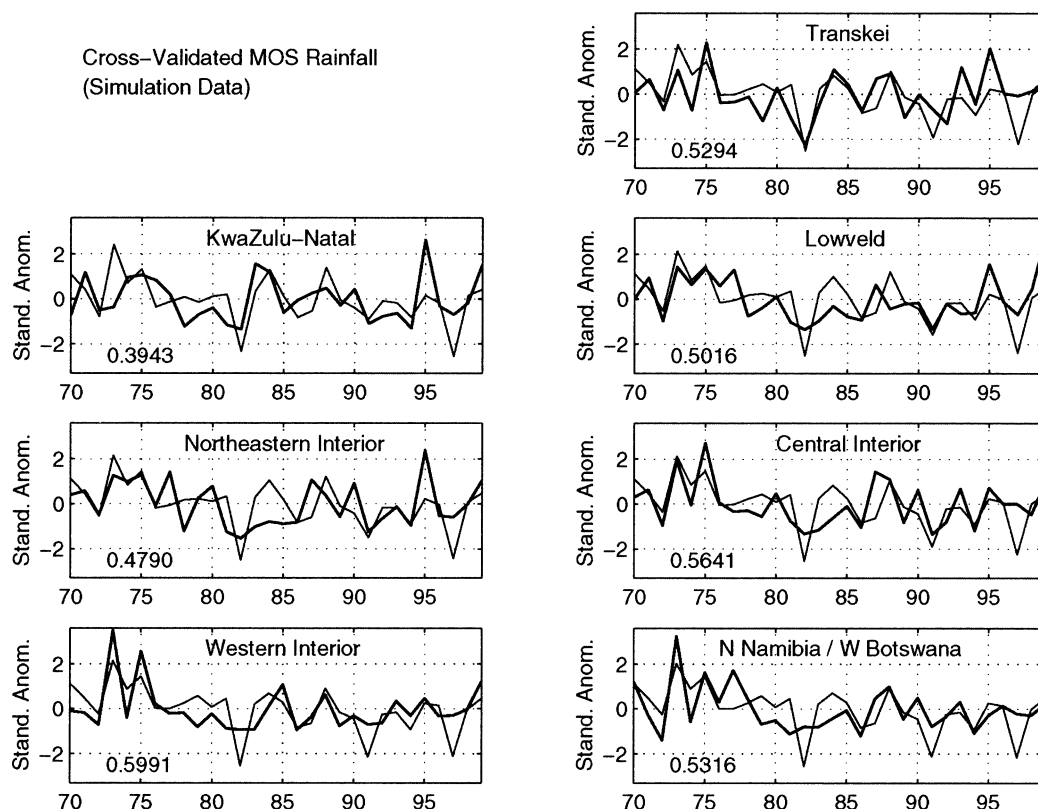


FIG. 3. Cross-validated simulation-MOS DJF rainfall standardized anomalies (thin line) vs the observed DJF rainfall standardized anomalies (thick line) for each of the seven summer rainfall regions. The correlation between the two time series of each region is in the bottom left-hand corner.

the observed rainfall anomaly during the weak La Niña event of 1995/96 is underestimated. In addition, the dry conditions predicted by both the MOS model and the area-averaged GCM rainfall data (Fig. 2) for one of the strongest El Niño events ever recorded, that of 1997/98, was not observed. It should be recognized that El Niño and La Niña events are not the only influence on southern Africa's climate, and also that seasonal climate in general is represented by a range of possibilities of which the observed reality constitutes only one of those possibilities.

In general, the simulation mode MOS model has been shown to outscore the GCM rainfall simulations (Fig. 4), justifying its use over the raw GCM rainfall. Large improvement in skill is found using the MOS model compared to the GCM over most of the regions during the 1990s. Although the improvements seem large qualitatively, few achieve statistical significance at the 95% level of confidence. Requiring such large improvements in skill makes it difficult to achieve strongly significant differences for fewer than the 30 years of training period considered here. Statistically significant correlation differences at the 95% level of confidence are, however, found for the KwaZulu-Natal region when the most recent 20–23 yr of the 30-yr training period are considered.

### c. Robustness of the predictor fields

The skillful rainfall simulations produced by the MOS model necessitates the investigation into the origin of its skill. Four different training periods are considered here in model validation: 30 yr of cross validation, and the assessment of operational forecast skill obtained from using training periods of 21, 24, and 27 yr, respectively. Each of these training periods starts in 1970/71. Although the best predictor field has been chosen using all the available data, it is assumed that the selected field would have been chosen regardless of the training period considered. Over the 30-yr period the 850-hPa geopotential height field is only slightly better than the other predictors, but they are all so highly correlated that even if a different predictor had proved slightly better over the shorter periods, the 850-hPa field would still remain an appropriate predictor. The optimal predictor–predictand mode combination producing the highest average correlation over the summer rainfall regions as discussed in section 4a, is assumed to also be the best estimate of the number of predictor and predictand modes used in the MOS equations for the four different climate periods. However, good skill is obtained with the optimal combination assumption for different periods only if the predictor–predictand setup is

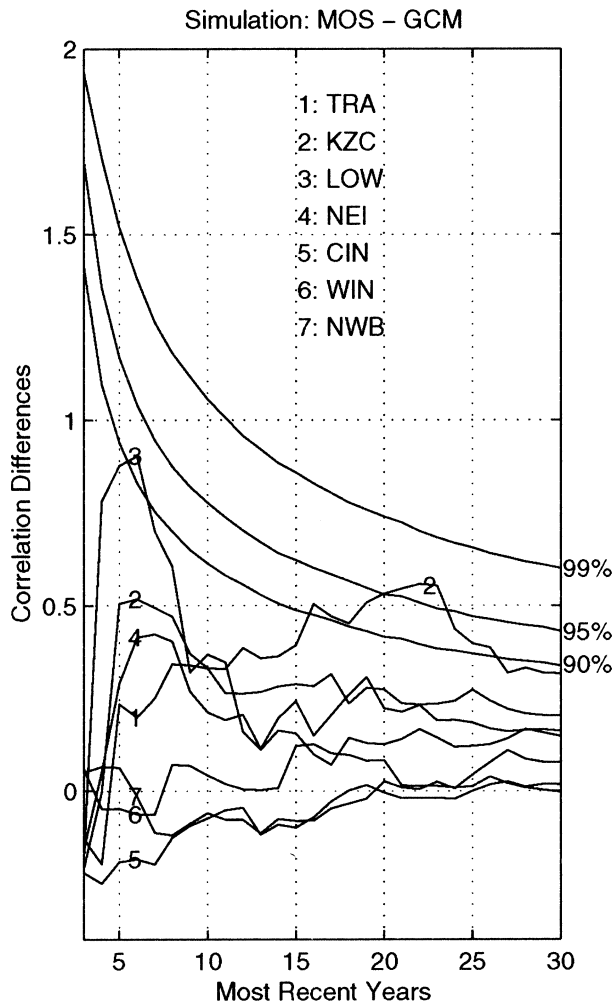


FIG. 4. Correlation differences between the MOS cross-validated simulations and the GCM-simulated area-averaged DJF rainfall for the seven summer rainfall regions (TRA: Transkei; KZC: KwaZulu-Natal; LOW: Lowveld; NEI: northeastern interior; CIN: central interior; WIN: western interior; NWB: northern Namibia/western Botswana). Leftmost correlation differences are associated with differences obtained by considering the most recent years of the 30-yr training period, and those on the right with the complete training period. The sloping lines indicate 90%, 95%, and 99% confidence levels.

robust. Given the high cross-validation correlations and assuming robustness, it should prove possible to achieve good forecast skill for the 9-yr independent retroactive predictions of DJF rainfall anomalies, regardless of the choice of using a 21-, 24-, or 27-yr training period.

Figure 5 shows the CCA diagnostic features, the predictor maps, the predictand maps, and their associated time series, behind the MOS recalibration linking the DJF 850-hPa *simulation mode* fields to the DJF regional rainfall anomalies. These maps and time scores (Barnett and Preisendorfer 1987; Landman and Mason 1999b; Landman and Tennant 2000) indicate the origin of the MOS model's forecast skill. In the discussion that follows, emphasis will be on CCA mode 1 because mode

2 provides only a small contribution to the predictability of the rainfall anomalies as reflected in the low canonical correlations of about 0.1 for each of the four training periods considered here. Although this low value might indicate that simple regression would perform just as well, there are a few years where the GCM-simulated 850-hPa geopotential height anomaly pattern closely resembles the CCA mode-2 spatial pattern, and has subsequently contributed to the MOS forecast skill.

For the first CCA mode anomalously low (high) GCM 850-hPa geopotential heights (Figs. 5a and 5b) over southern Africa are associated with rainfall above (below) the average over most of the region (Figs. 5c and 5d). This association is emphasized by the high correlation values between the predictor and predictand CCA time scores (Figs. 5e and 5f). An area of anomalously low (high) GCM 850-hPa geopotential heights over the central and western subcontinent and an associated anomalously high (low) 850-hPa geopotential height area south and southeast of the continent are supported by analysis based on observations of rain (drought) producing systems (Tyson and Preston-Whyte 2000), evidence that the observed large-scale structure and variability are well simulated by the GCM. Little difference is found in the associations in either the spatial correlations of the predictor and predictand or the time scores over the training periods considered. Although Fig. 5 only shows the analyses of the 21- and 27-yr training periods, almost identical features are also found for the 24- and 30-yr periods. In addition, similar robustness is also found for mode 2. This stationarity demonstrates the robustness of the predictor–predictand fields regardless of the length of the training sets considered here, which justifies the assumption that the statistical relationships between the GCM-derived large scale and the observed regional rainfall are valid at least during the independent test period.

The simulation mode MOS rainfall holds no operational forecast value because the GCM is forced with SST anomalies that are simultaneous with the rainfall training season. To judge forecast value, the MOS model's performance must be assessed in a forecast setting. In hindcast mode, the GCM is forced by persisting November SST anomalies through the DJF forecast season, thus imposing a lead time. Albeit short, this lead time is in effect associated with a real-time forecast issued in early December for the DJF season, and is referred to here as a 1-month lead time. The CCA diagnostic features of the DJF 850-hPa geopotential height *hindcast mode* field and the observed rainfall are shown in Fig. 6. Similar features producing wet and dry conditions respectively for the four training periods considered are found using the hindcast mode data: strong correspondence is again found for the predictor correlation maps (Figs. 5a,b and 6a,b), the predictand map correlations for both the simulation and hindcast mode cases (Figs. 5c,d and 6c,d), and the canonical time scores (Figs. 5e,f and 6e,f). These results confirm that persisted SST anomalies are

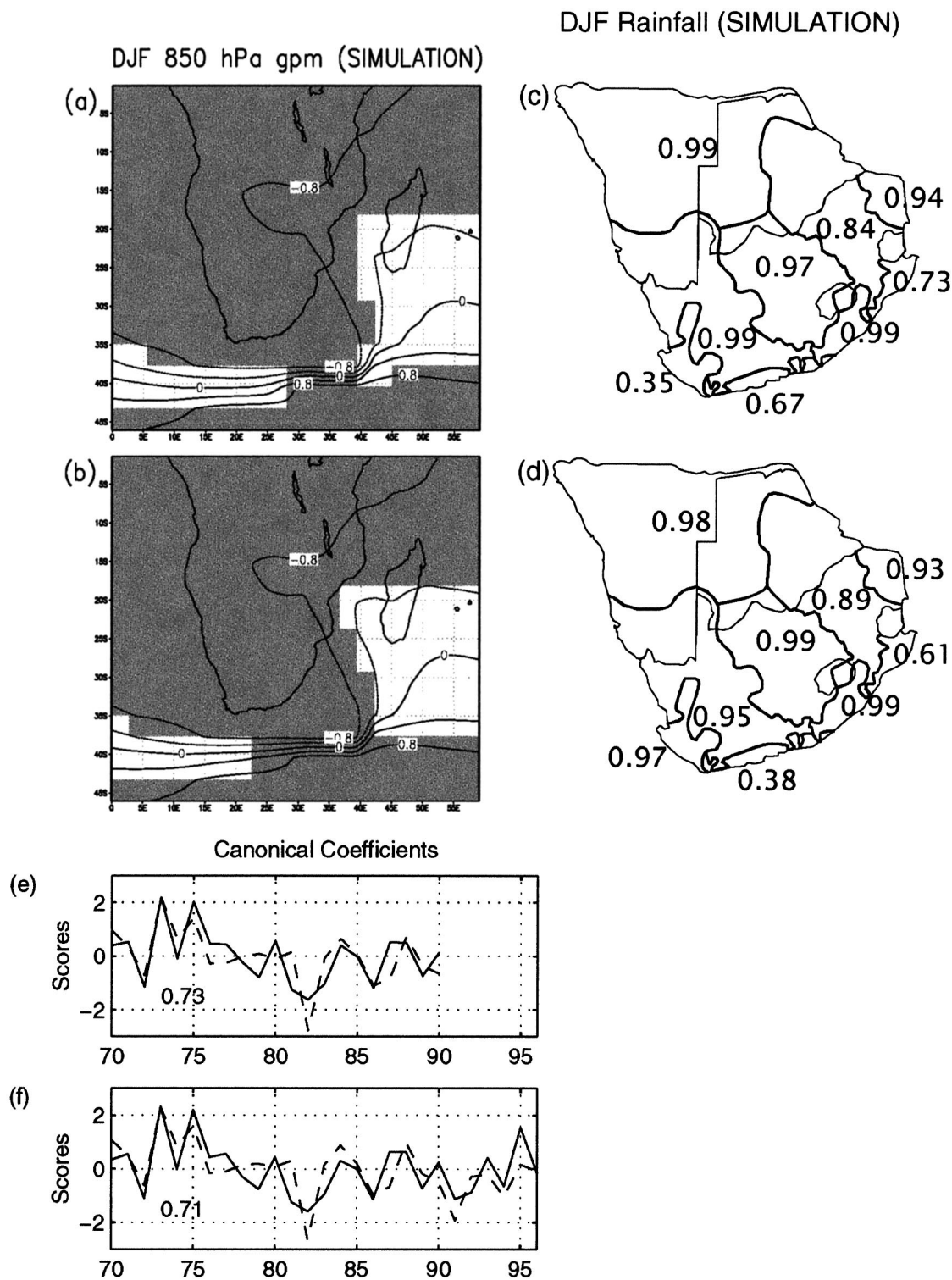


FIG. 5. CCA mode-1 predictor maps of the (a) 21- and (b) 27-yr training periods used in the MOS equations relating simulation mode 850-hPa geopotential heights to DJF regional rainfall. Shaded regions depict areas of significant correlations at the 95% level of confidence. CCA mode-1 predictand maps of the (c) 21- and (d) 27-yr training periods used in the MOS equations. Canonical coefficient scores of mode 1 of the (e) 21- and (f) 27-yr training periods, of the 850-hPa geopotential heights (dashed line), and the regional rainfall (solid line); canonical correlations are in the bottom left-hand corner.



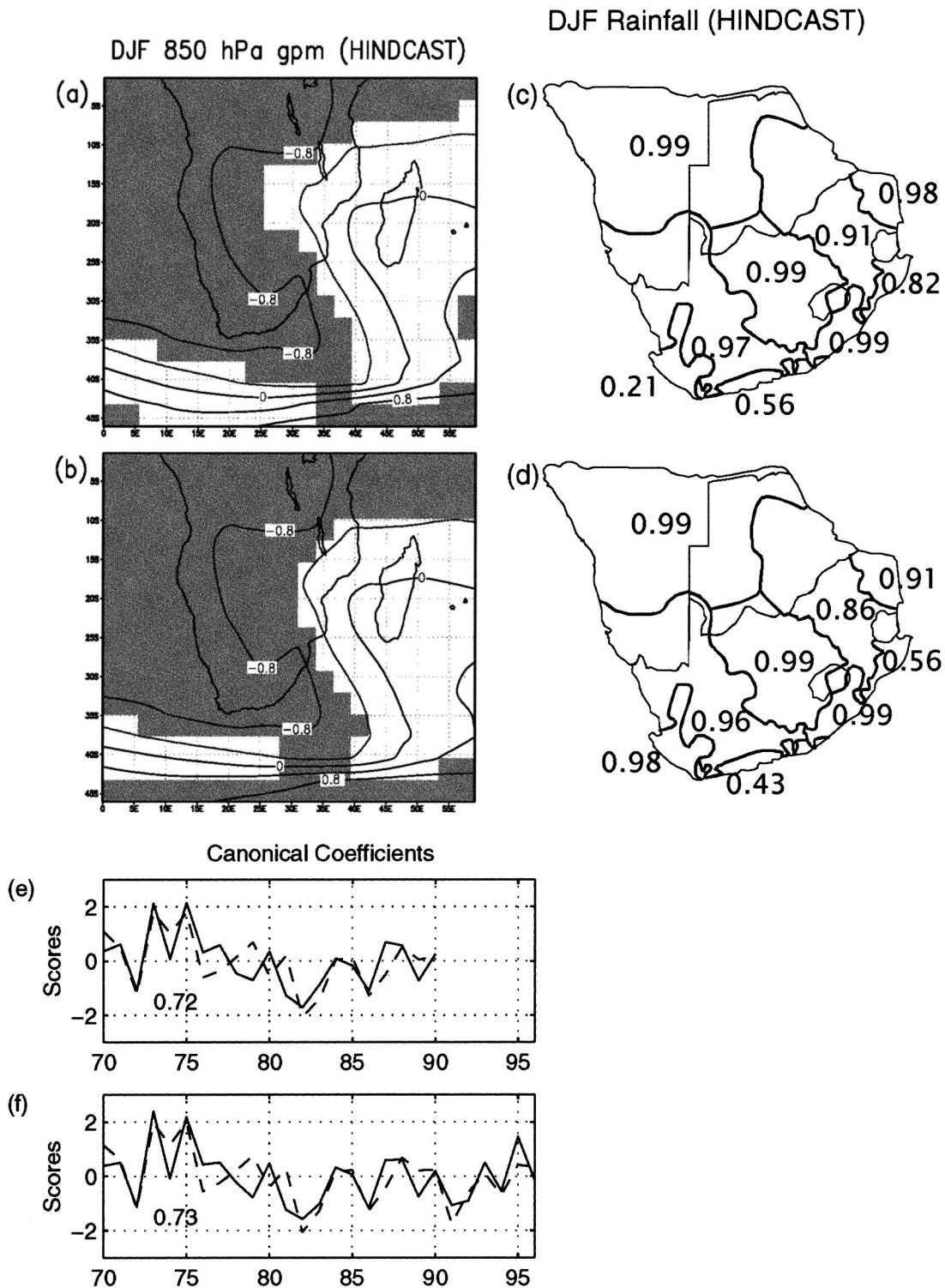


FIG. 6. As in Fig. 5, but for hindcast data.

a sufficient representation of the actual SST anomalies, thus justifying the use of the hindcast 850-hPa geopotential heights in the MOS model to produce forecasts at the proposed lead time. In addition, the similarities

suggests that similarly skillful MOS rainfall predictions may be produced using hindcast mode GCM fields as are found using simulation mode GCM fields as predictor in the MOS equations.

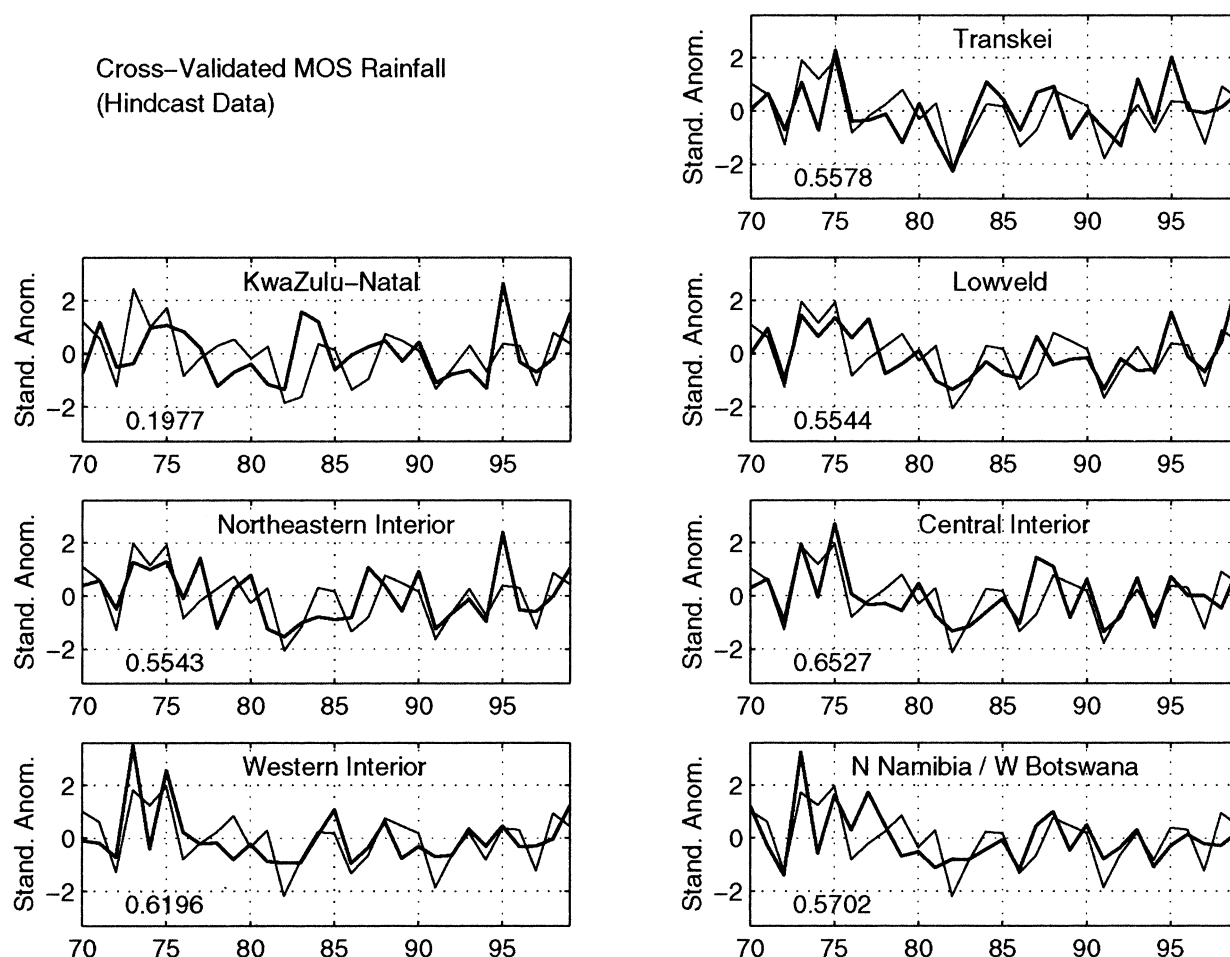


FIG. 7. Cross-validated hindcast-MOS DJF rainfall standardized anomalies (thin line) vs the observed rainfall standardized anomalies (thick line) for each of the seven summer rainfall regions. The correlation between the two time series of each region is in the bottom left-hand corner.

#### d. Performance of the MOS model using hindcast data

The cross-validated hindcast rainfall MOS forecasts are presented in Fig. 7. The hindcast MOS forecasts are associated with typically higher cross-validation correlation values than those found for the simulation data (Fig. 3). The differences, however, are not statistically significant. The important result is that in most cases the skill of the MOS model did not deteriorate when applied at the lead time of 1 month. Close inspection of both the cross-validation simulation-MOS forecasts (Fig. 3) and the hindcast-MOS forecasts (Fig. 7) reveals a downward trend in MOS rainfall anomalies. A similar trend is found in the dominant predictor canonical time scores as well (Figs. 5e,f and 6e,f). However, a Monte Carlo significance test showed that the trends' significance associated with the hindcast-MOS data lies between the 77% and 80% confidence levels, making it unlikely that operational forecast skills are coming about largely due to trends.

It has been determined that the MOS model does improve on the GCM rainfall fields and that the MOS model remains skillful at a forecast lead time. The next question is whether the forecast system that uses a MOS recalibration can outscore a much simpler and a computationally less expensive linear model of southern African seasonal rainfall. A CCA rainfall forecast model (Landman and Mason 1999b) is utilized to relate September–November (SON) SST anomalies with observed DJF rainfall over the 30-yr climate period of 1970/71–1999/2000 (Fig. 8). An optimal skill model is designed similarly to the MOS recalibration of the GCM using cross-validation sensitivity runs.

For the empirical rainfall forecasts using SON SSTs as predictor, all of the cross-validation correlation values are statistically significant at the 95% level of confidence. Forecast during the 1990s are skillful, except for the wet conditions predicted for the La Niña season of 1998/99 and, as is found for the MOS forecast system, inaccurate rainfall forecast during the 1997/98 El Niño

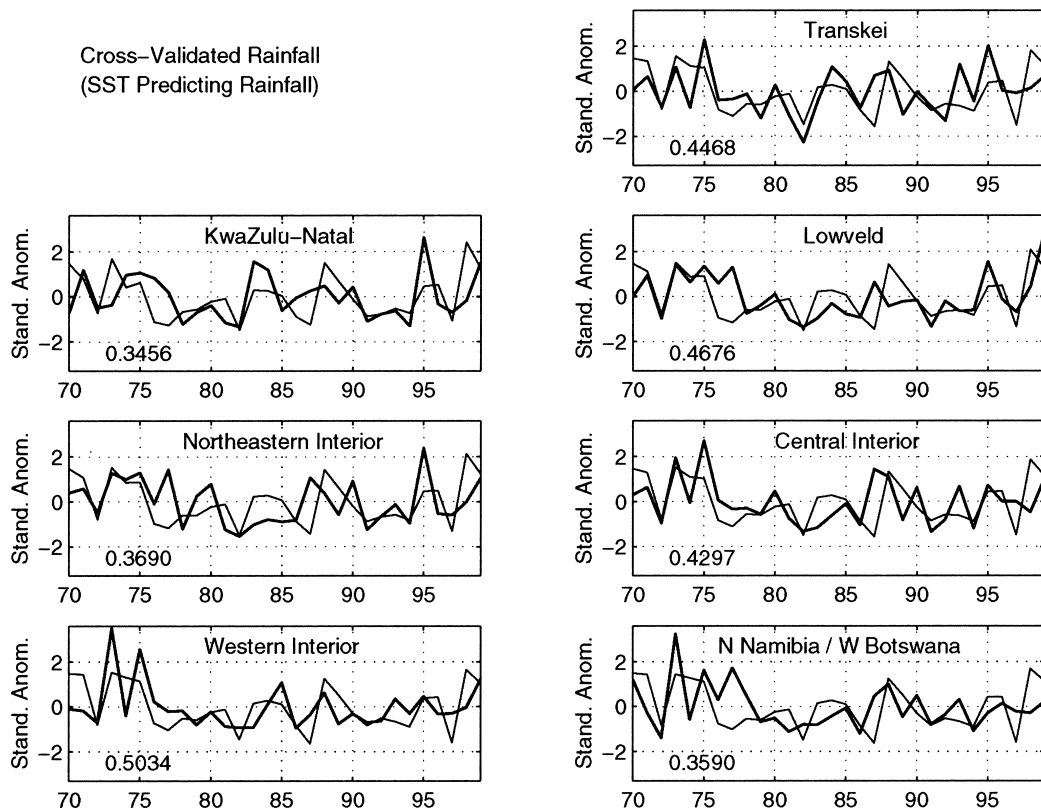


FIG. 8. Cross-validated DJF rainfall standardized anomalies (thin line), using SON SSTs as predictor, vs the observed DJF standardized rainfall anomalies (thick line) for each of the seven summer rainfall regions. The correlation between the two time series of each region is in the bottom left-hand corner.

event. In general, the hindcast-MOS skill over the cross-validation period is possibly better than that of the linear empirical forecast model since positive differences are found for most of the regions (Fig. 9). The MOS model has thus outscored both of the baseline skill levels: the GCM area-averaged rainfall and a simple linear model.

Although the correlation values of the MOS model may suggest a skillful rainfall forecast scheme, cross validation may still indicate biased skill levels (Barnston and van den Dool 1993). The MOS model's operational forecast skill will subsequently be conducted over a test period that is independent of any training period.

#### e. Categorized real-time forecasts

In this section, the hindcast-MOS model's categorized rainfall forecasts are validated over a retroactive forecast period of 9 yr (1991/92–1999/2000), and also over the 30-yr cross-validation period. The assessment of the MOS model's performance over the retroactive period provides a rigorous test of its ability to produce useful rainfall forecasts in a real-time operational forecast environment, since the model is trained only on data prior to the target years of the retroactive period. At each point, or region, the time series for each of the MOS and observed fields is separated into its own three equi-

probable groups defining above-normal, near-normal, or below-normal conditions. The categorized recalibrated forecasts are compared with that of the observed in order to calculate the categorical skill, expressed as LEPS scores, of the forecast system.

Significant LEPS scores are found for both the cross-validation period (Fig. 10a) as well as the retroactive period (Fig. 10b). Most of the LEPS scores for both the 30- and 9-yr period are statistically significant at the 95% level of confidence. However, the LEPS scores of the 9-yr retroactive period are found to be much higher than those of the 30-yr cross-validation period. The greatest increases in LEPS scores are found for KwaZulu-Natal (63.4), followed by the Lowveld (38.0), the northeastern interior (33.2), northern Namibia/western Botswana (23.6), the central interior (23.4), and Transkei (16.9). The higher scores obtained during the retroactive period are further illustrated by investigating the skill levels of three consecutive 9-yr periods contained within the 30-yr cross-validation period, that is, 1973/74–1981/82, 1982/83–1990/91, and 1991/92–1999/2000. For most of the regions, both the LEPS scores (Fig. 11a) and correlation values (Fig. 11b) obtained from the three 9-yr periods contained within the 30-yr period are the highest during the retroactive 9 years, and again the largest differences are found for

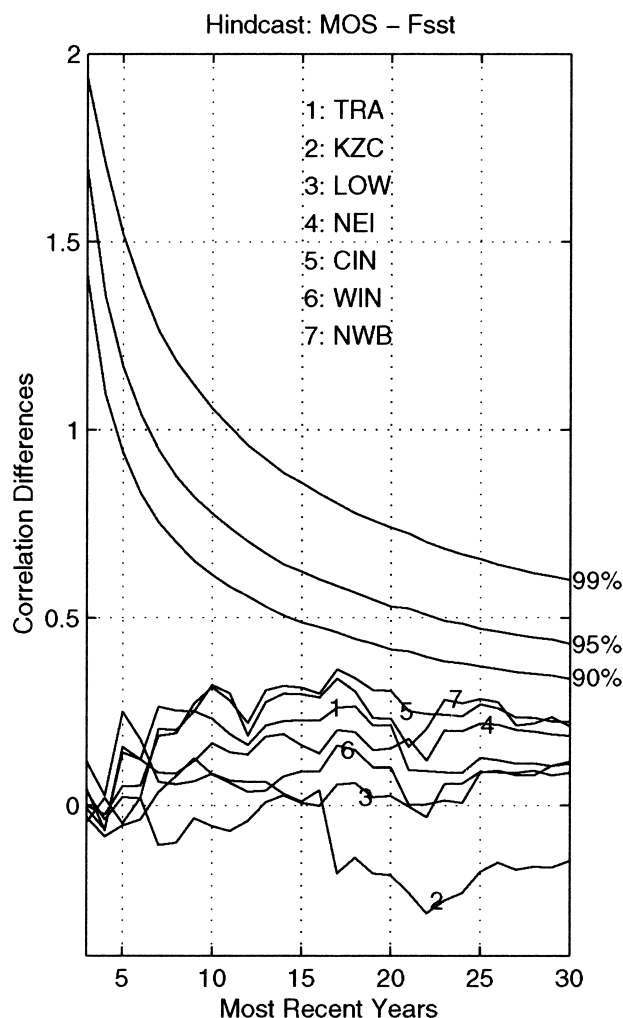


FIG. 9. Correlation differences between the MOS cross-validated hindcasts and the statistical model rainfall forecasts of DJF rainfall for the seven summer rainfall regions (for region definitions, see Fig. 4). Leftmost correlation differences are associated with differences obtained by considering the most recent years of the 30-yr training period, and those on the right with the complete training period. The sloping lines indicate 90%, 95%, and 99% confidence levels.

KwaZulu-Natal, followed by the Lowveld and northeastern interior. The reason why the LEPS scores obtained from the retroactive predictions (dashed line in Fig. 11a) differ from that of the most recent 9-yr period of the cross-validation period (solid line with asterisks in Fig. 11a) is because the three equiprobable category boundaries associated with the retroactive process are somewhat different for each of the three different climate periods of, respectively, 21, 24, and 27 yr involved in training the retroactive MOS model. They are also somewhat different from the categories associated with the 30-yr cross-validation period, which will ultimately affect the number of category hits and consequently the LEPS scores.

#### f. Probability forecasts

Since seasonal climate is inherently probabilistic, seasonal forecasts should be judged probabilistically. One measure is to quantify the relative confidence of forecasting the observed category using a metric such as the ranked probability skill score (RPSS). The ranked probability score (RPS), is an extension of the Brier score (Wilks 1995) to the multievent situation, while the RPSS is a skill score for a collection of RPS values relative to the RPS obtained from climatological probabilities (Wilks 1995). A collection of perfect forecasts, predicting the observed category with 100% probability, would have an RPSS of 1, and a perpetual forecast of climatological probabilities would yield a RPSS of 0. The 1991/92–1999/2000 9-yr retroactive period has been shown to be associated with high skill for the deterministic and categorical presentations of the rainfall forecasts. The operational utility of the hindcast-MOS model is further assessed by calculating the RPSS from the probability forecasts generated by MOS forecasts consisting of five ensemble members. The forecast skills based on the probability forecasts of the retroactive period are similar to those found for the LEPS scores of the deterministic retroactive forecasts (Fig. 10), with the highest RPSS found for the KwaZulu-Natal region and the central interior followed by the RPSS of the Lowveld and northeastern interior regions (Fig. 12).

Probabilities are assigned to each category by calculating the number of times a category is hit by any of the five ensemble members. For example, if 2 of the 5 ensemble members are in the near-normal category and the remainder in the above-normal category, then the probability forecast for that season is for a 0% chance of below-normal rainfall, a 40% chance of near-normal rainfall, and a 60% chance of above-normal rainfall. Clearly having only five ensemble members from which to produce a probabilistic forecast involves sampling errors, but it does give some indication of possibly forecast spread or uncertainty. Table 1 represents the deterministic and probabilistic forecasts for the 9-yr retroactive test period. An example of the advantage of using probabilistic forecasts instead of deterministic forecasts is demonstrated by the rainfall forecasts of the 1996/97 season presented in the table: the deterministic forecasts are above normal for most regions while the observed categories for all the regions are near normal, which is in agreement with the largest probability category. In this case, the deterministic forecasts falls in one category even though the largest probability of occurrence is in another because one or two of the ensemble members have larger absolute values than the rest of the members forcing the ensemble mean toward the outlying value. In addition to the demonstrated value of the probabilistic forecasts of the 1996/97 season, the probability forecasts of the 1997/98 season can also be viewed as an improvement over the deterministic forecast of below normal because some probability was as-



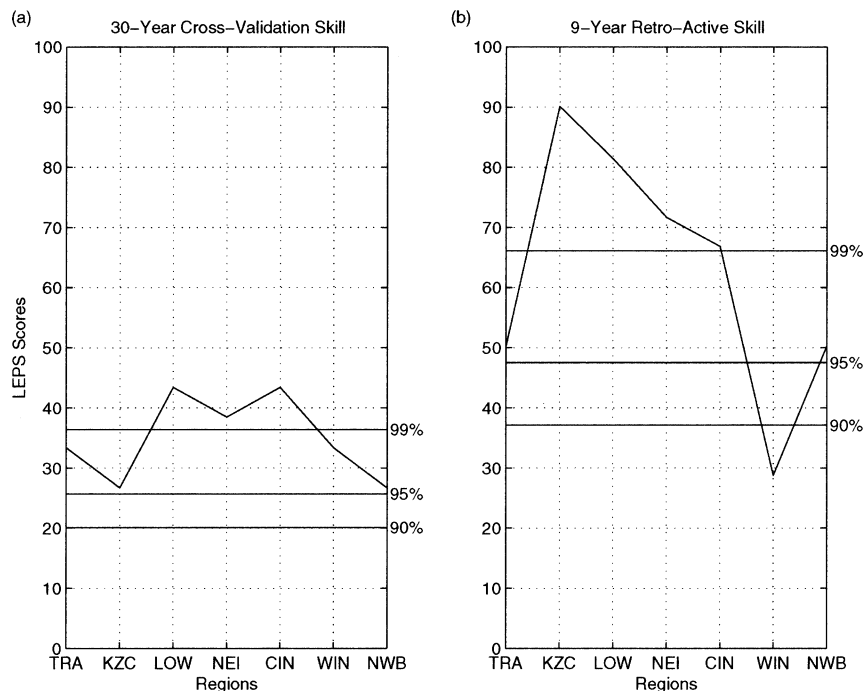


FIG. 10. LEPS scores for (a) the 30-yr cross-validation period (1970/71–1999/2000) and (b) the 9-yr retroactive forecast period (1991/92–1999/2000) for the seven summer rainfall regions (for region definitions, see Fig. 4). The horizontal lines indicate 90%, 95%, and 99% confidence levels.

signed to the near-normal category, which is the observed category for most of the regions. Notwithstanding the improvement in the value of the 1997/98 forecast through a probability distribution, the fact that only one ensemble member indicated a near-normal category for this season while the remaining four indicated a below-normal category is an example of a situation where nature may not have resulted in the most likely outcome. In addition, the probability rainfall forecasts of the 1991/92 El Niño and the 1999/2000 La Niña events did not add value to the deterministic forecast of mostly below- and above-normal rainfall, respectively, manifested by the agreement between the deterministic forecast category, the probability forecast of 100%, and the observed rainfall category. Furthermore, some of the 1994/95 probability forecasts, for example, did not add value either, owing to the most likely category being one category out.

#### g. Streamflow forecasts with hindcast GCM fields

Atmospheric GCMs do not explicitly simulate streamflow, necessitating the statistical link between GCM-simulated fields and streamflow. In addition, the GCM used here has a much coarser resolution than the distances between the inlets of the dams. Thus recalibrating the GCM output to streamflow is truly a downscaling exercise. The recalibration procedure using hindcast data for forecasting rainfall is next applied to

the streamflow at the inlets of six dams in the Vaal and upper Tugela River catchments, which lie within the northeastern interior region. Only the cross-validated forecasts are presented for the period 1971/72–1994/95. The naturalized streamflow data used in this paper are not available for the period after early 1995. The same predictor set, the hindcast mode 850-hPa geopotential height field that is used to recalibrate to seasonal rainfall anomalies is used here, because streamflow is directly affected by precipitation and its variability should therefore similarly be affected by the variability of the 850-hPa geopotential heights.

As before, sensitivity runs using cross validation are performed to obtain the optimal streamflow downscaling model. Using three predictand and five predictor modes in the model produced the highest averaged cross-validation correlation value, with each set of modes explaining more than 90% of the respective total variances. Additional factors affecting streamflow are evaporation and changes in soil moisture, as well as nonmeteorological factors such as vegetation cover and the soil surface characteristics of catchments. The association between rainfall and streamflow is therefore complex, and also depends on factors that are not directly related to atmospheric variability. However, none of these factors are explicitly simulated by the atmospheric GCM and thus cannot be incorporated into the downscaling process described in this paper. This downscaling model, however, can at least set a baseline

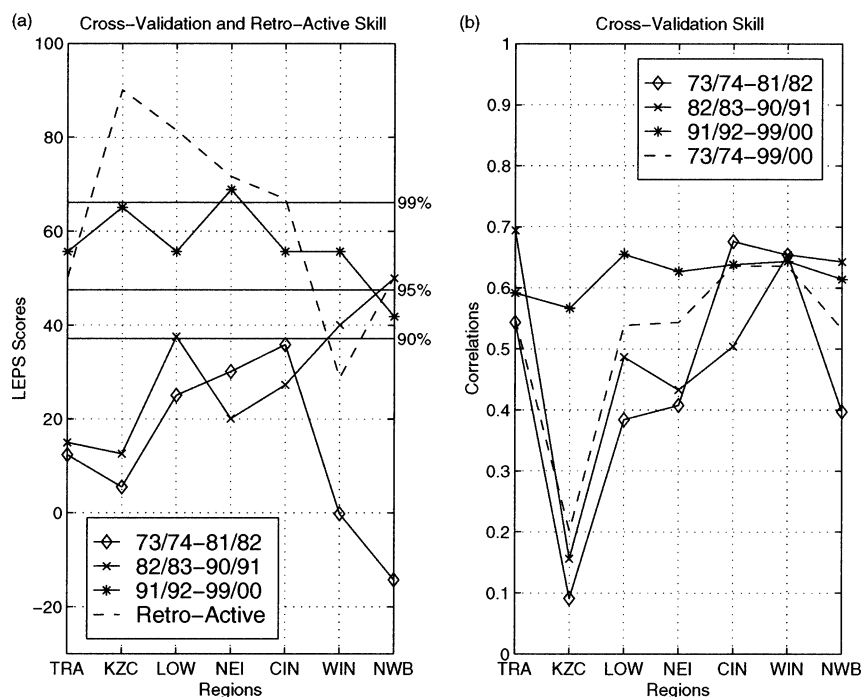


FIG. 11. (a) Cross-validation LEPS scores for three consecutive 9-yr periods within the 30-yr training period, and LEPS scores for the 9-yr retroactive forecast period (dashed line). The horizontal lines indicate 90%, 95%, and 99% confidence levels. (b) Cross-validation correlations for three consecutive 9-yr periods within the 30-yr training period and also for the most recent 27 yr in the training period. Scores are shown for the seven summer rainfall regions (for region definitions, see Fig. 4).

against which other more complex downscaling processes can be compared.

The main purpose of this section is to assess if the proposed MOS can be of some value as an operational applications forecast procedure. Cross validation is performed on each of the five hindcast ensemble members and the average of the forecasts is obtained. The cor-

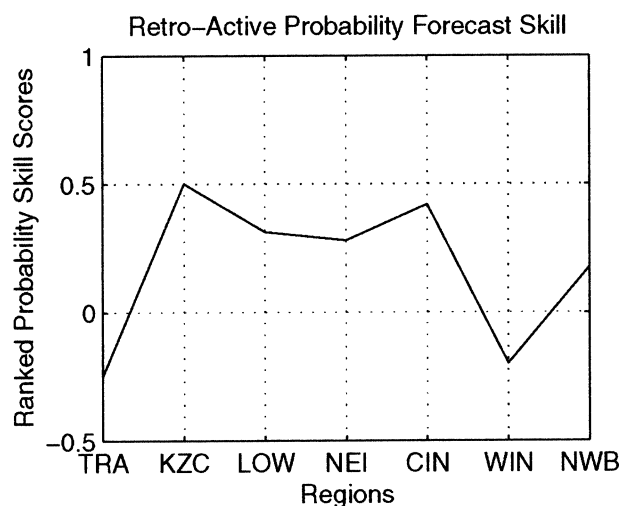


FIG. 12. RPSS of the 9-yr retroactive forecast period for the seven summer rainfall regions (for region definitions, see Fig. 4).

relation values between the ensemble mean MOS and the observed streamflow vary between 0.54 for the Vaal Dam and 0.65 for the Johan Nesser Dam (Fig. 13). A high association is found between the observed streamflow and the observed rainfall of the region that contains the catchments of the dams. The high association is a manifestation of the effect rainfall has on the streamflow at the inlets of these dams, and is proof that the 850-hPa geopotential height field that contributed to the rainfall prediction skill is a reasonable choice as predictor for streamflow also. Streamflow forecast skill should improve further if other nonatmospheric variables were allowed to participate in the recalibration process. As is the case in the rainfall recalibration, improved streamflow forecasts also occurred after the 1989/90 season.

## 5. Summary and discussion

A model output statistics (MOS) method to statistically recalibrate the circulation patterns generated by an atmospheric GCM has been presented for the DJF rainfall season over southern Africa. The best estimate of the GCM prognostic fields, that is, GCM-generated fields produced from forcing the GCM with observed SST anomalies, was first used to assess if simulation-MOS "forecasts" produced from the recalibration can

TABLE 1. Deterministic and probabilistic 1-month lead real-time DJF rainfall MOS forecasts for the 9-yr retroactive test period of 1991/92–1999/2000. Two columns are seen at each season of which the left-hand column represents the probability distribution of the forecast based on the five hindcast ensemble members; the right-hand column shows the deterministic forecast and also the observed rainfall category. Here “A” and “a” refer to above normal; “N” and “n” refer to near normal; “B” and “b” refer to below normal. Forecast categories are depicted by lower case and observed categories by upper case in each of the right-hand columns; bold entries are associated with a deterministic forecast hit.

	1991/92	1992/93	1993/94	1994/95	1995/96	1996/97	1997/98	1998/99	1999/2000
<b>Transkei</b>									
A	0	—	0.4 A	0	0.2 Aa	0.2 —	0	0.6 a	1 Aa
N	0	0.6 —	0.6 n	0.2 N	0.6 —	0.6 Nn	0.2 N	0.2 N	0 —
B	1 b	0.4 Bb	0	0.8 b	0.2 —	0.2 —	0.8 b	0.2 —	0 —
<b>KwaZulu-Natal Coast</b>									
A	0	—	0.8 —	0.2 —	0.6 Aa	0.4 a	0	0.4 a	0.8 Aa
N	0.2 —	0.4 —	0 —	0 —	0.4 —	0.4 N	0	0.4 N	—
B	0.8 Bb	0.6 Bb	0.2 Bb	0.8 Bb	0 —	0.2 —	1 Bb	0.2 —	0 —
<b>Lowveld</b>									
A	0	—	0.6 —	0 —	0.2 Aa	0.2 a	0	0.4 Aa	1 Aa
N	0	0.4 N	0.4 Nn	0.2 N	0.8 —	0.6 N	0.2 —	0.4 —	0 —
B	1 Bb	0.6 b	0 —	0.8 b	0 —	0.2 —	0.8 Bb	0.2 —	0 —
<b>Northeastern interior</b>									
A	0	—	0.6 —	0 —	0.2 Aa	0.2 a	0	0.4 a	0.8 Aa
N	0	0.4 N	0.2 Nn	0.2 —	0.8 —	0.6 N	0.2 N	0.4 N	—
B	1 Bb	0.6 b	0.2 —	0.8 Bb	0 —	0.2 —	0.8 b	0.2 —	0 —
<b>Central interior</b>									
A	0	—	0.4 A	0 —	0.2 Aa	0.2 a	0	0.6 a	1 Aa
N	0	0.4 —	0.6 n	0.2 —	0.8 —	0.6 N	0.2 N	0.2 N	—
B	1 Bb	0.6 Bb	0 —	0.8 Bb	0 —	0.2 —	0.8 b	0.2 —	0 —
<b>Western interior</b>									
A	0	—	0.2 A	0 —	0.2 A	0.2 —	0	0.6 Aa	1 Aa
N	0	0.6 N	0.8 n	0.2 N	0.8 N	0.6 Nn	0.2 N	0.2 —	0 —
B	1 b	0.4 b	0 —	0.8 b	0 —	0.2 —	0.8 b	0.2 —	0 —
<b>Northern Namibia/western Botswana</b>									
A	0	—	0.4 —	0 —	0.2 a	0.2 —	0	0.6 a	0.8 a
N	0	0.6 N	0.4 Nn	0.2 —	0.8 N	0.6 Nn	0.2 N	0.2 N	0.2 N
B	1 Bb	0.4 b	0.2 —	0.8 Bb	0 —	0.2 —	0.8 b	0.2 —	0 —

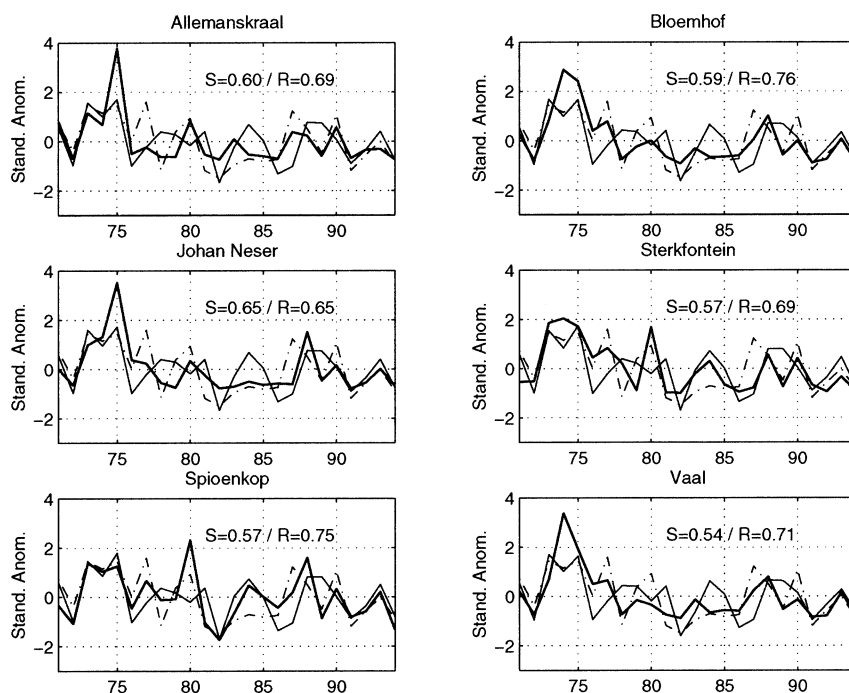


FIG. 13. Cross-validated hindcast-MOS DJF streamflow standardized anomalies (thin line) vs the observed DJF streamflow standardized anomalies (thick line) for each of six dams of the Vaal and upper Tugela River catchments. DJF rainfall standardized anomalies (dashed-dotted line) of the northeastern interior are also shown. The correlations between the predicted and observed streamflow standardized anomalies ( $S$ ) and the observed streamflow and rainfall standardized anomalies ( $R$ ) are shown in the top right of each panel.

outscore the simulation rainfall data generated by the GCM. The origin of the MOS model's skill was investigated next to see if the associated patterns of atmospheric circulation and rainfall variability can be supported by observations. However, the simulation-MOS data do not hold any operational forecasting utility owing to the simultaneity of the SST forcing field and the rainfall season. Hindcast data, produced by forcing the GCM with persisted November SST anomalies through DJG, provide the required input to the MOS equations that can ultimately produce real-time forecasts at the beginning of December for the DJF rainfall season. The operational forecast ability of the hindcast-MOS model is further assessed through a retroactive forecast scheme. Finally, the ability to use the MOS model in a streamflow applications experiment is demonstrated.

After testing an array of possible candidate GCM-generated predictor fields the 850-hPa geopotential height field is chosen as the only predictor in the MOS set of equations due to its more skillful results than that produced with other candidate predictor fields. The MOS model improved on the area-averaged GCM rainfall, which is particularly evident during the most recent decade of the 30-yr training period. Spatially, the improvement is significant over the northeastern interior of South Africa, which may be attributed to the spatial correction of the seasonal average GCM rainfall distribution: although the GCM is successful in simulating

the overall pattern of maximum rainfall over the northeast decreasing toward the southwest, that is, from the Lowveld toward the southwestern Cape (see Fig. 1), it has displaced slightly the local maximum over these regions.

Both simulation- and hindcast-MOS forecasts show significant agreement with the observations. This agreement is especially evident during strong El Niño–Southern Oscillation (ENSO) events when accurate forecasts are produced for mostly above-average rainfall during La Niña events and below-average rainfall during El Niño events. The dominant canonical times scores additionally suggest a strong ENSO relationship with the rainfall anomalies, in such a way that anomalously low (high) 850-hPa geopotential heights are observed over the central and western interior regions of southern Africa when the DJF rainfall is found to be above (below) the average. Observations support this physical association between the GCM-produced height field and the rainfall, which is found to be robust over four different model training periods. The similarities in the analyses is evidence that the large-scale structure and variability of the 850-hPa geopotential height field are well characterized by the GCM regardless of the two SST forcing fields used here, and also why equally skillful MOS rainfall predictions are produced using either hindcast and simulation mode GCM fields as predictor in the MOS equations. Thus, real-time MOS DJF rainfall fore-



casts produced at the beginning of December are as skillful as one can expect to find with this forecast system, but has the added advantage that these forecasts have the potential to aid in decision making in an operational environment.

The hindcast-MOS model is further tested against an optimal simple linear statistical model that makes use of global SON SSTs as predictor of DJF rainfall. These forecasts thus have the same lead time as that of the hindcast-MOS model and produced similar skill levels, which demonstrates how hard it is to beat simple linear statistical models of seasonal rainfall especially in the absence of dominating nonlinearities in the atmosphere-ocean system (Kumar and Hoerling 2000). However, the SST-rainfall associations between southern African summer rainfall and equatorial Indian Ocean temperatures is unstable, but the instability will be best simulated using GCMs (Landman and Mason 1999a). Therefore, most of the skill differences between the hindcast MOS and the simple linear statistical model are expected and found to be positive for most cases. The MOS has demonstrated its potential to outscore GCM-generated rainfall forecasts and also forecasts from an inexpensive, simple linear model, therefore justifying its use in an operational forecast setting.

As categorized forecasts, particularly tercile forecasts, are becoming a standard format in seasonal forecasting, the MOS model's rainfall forecasts are categorized and the associated skill assessed for the most recent 9 yr in the 30-yr dataset through a retroactive forecast procedure. Categorized forecasts are also produced for the entire 30-yr training period and its skill levels compared with that of the retroactive 9-yr period. Skill levels of the retroactive period are mostly much larger than that of the 30-yr period, especially for the northeastern interior regions of South Africa. The reason why these regions have experienced such a large increase in prediction skill during the retroactive years is not obvious, but small samples of about 10 yr within the 30-yr period could be associated with varying skill levels as has also been found in the interdecadal predictability of ENSO events (Trenberth and Hurrell 1994). In addition, owing to the strong ENSO signal in the predictor-predictand setup, the higher number of observed ENSO events during the most recent decade of the 30-yr period will also lead to improved forecast skill.

Additional forecast value is obtained in knowing the probability of a predicted category occurring (Mason and Graham 1999). In fact, probability forecasts exhibit reliability considerably in excess of that achieved by corresponding nonprobabilistic forecasts (Murphy 1998). The ensemble of five members used here, however, will provide only crude probability estimates, since variations in the forecasts could result as much from sampling variations in these small samples as from real forecast performance differences. Notwithstanding, the value of probability forecasts has been demonstrated

through the examples of the 1996/97 and 1997/98 rainfall forecasts. In contrast, however, one can point to examples within the retroactive period that did not add any additional forecast value.

Streamflow, which is directly affected by rainfall, justifies the use of the same predictor field that simulated seasonal rainfall successfully. Thus the MOS model is tested for streamflow forecasts at the inlets of six dams within the Vaal and upper Tugela River catchments. Although streamflow is dependent on factors other than rainfall that are not directly related to atmospheric variability, none of these factors are explicitly simulated by the atmospheric GCM and therefore could not be incorporated in the MOS model presented here. Notwithstanding, the potential value of the forecast system as an applications tool is demonstrated by its significantly high skill values, which should improve further if other nonatmospheric variables such as soil moisture and groundwater storage of the catchment are allowed to participate in the recalibration process.

The paper has demonstrated the efficacy of statistical interpretation of GCM output, or more specifically, recalibration of GCM output necessary to skillfully simulate seasonal rainfall and streamflow over southern Africa at a lead time that could be beneficial to the users of such forecast products. The MOS recalibration has shown improved skill over both the GCM-simulated rainfall as well as over a simple statistical forecasting technique. The MOS recalibration has also demonstrated, through real operational forecast examples, the added value that could potentially be obtained if such forecasts are used in a probabilistic sense.

*Acknowledgments.* This postdoctoral research, administered by the University Corporation for Atmospheric Research, was supported by cooperative agreement NA07GP0213 between the National Oceanic and Atmospheric Administration and Columbia University. Discussions with, and comments and suggestions by Tony Barnston are gratefully acknowledged. Also appreciated are the comments of the two anonymous referees of this paper. Anna Bartman made some of the figures.

## REFERENCES

- Barnett, T. P., and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States air temperature determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850.
- Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513–1564.
- , and H. M. van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate*, **6**, 963–977.
- , and T. M. Smith, 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Climate*, **9**, 2660–2697.
- Busuioac, A., C. Deliang, and C. Hellström, 2001: Performance of statistical downscaling models in GCM validation and regional

- climate change estimates: Application for Swedish precipitation. *Int. J. Climatol.*, **21**, 557–578.
- Cane, M. A., G. Eshel, and R. W. Buckland, 1994: Forecasting Zimbabwean maize yield using eastern Pacific sea surface temperatures. *Nature*, **370**, 204–206.
- Crane, R. G., and B. C. Hewitson, 1998: Doubled CO<sub>2</sub> precipitation changes for the Susquehanna Basin: Downscaling from the GEN-ESIS General Circulation Model. *Int. J. Climatol.*, **18**, 65–76.
- Cui, M., H. von Storch, and E. Zorita, 1995: Coastal sea level and the large-scale climate state. A downscaling exercise for Japanese Islands. *Tellus*, **47A**, 132–144.
- Deutsches Klimarechenzentrum, 1992: The ECHAM3 Atmospheric General Circulation Model. Tech. Rep. 6, DKRZ, Hamburg, Germany, 184 pp.
- Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624.
- Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152.
- Hastenrath, S., L. Greischar, and J. van Heerden, 1995: Prediction of summer rainfall over South Africa. *J. Climate*, **8**, 1511–1518.
- Hunt, B. G., 1997: Prospects and problems for multi-seasonal predictions: Some issues arising from a study of 1992. *Int. J. Climatol.*, **17**, 137–154.
- Jackson, J. E., 1991: *A User's Guide to Principal Components*. Wiley, 569 pp.
- Ji, M., A. Kumar, and A. Leetmaa, 1994: A multiseasonal climate forecast system at the National Meteorological Center. *Bull. Amer. Meteor. Soc.*, **75**, 569–577.
- Joubert, A. M., and B. C. Hewitson, 1997: Simulating present and future climates of southern African using general circulation models. *Prog. Phys. Geogr.*, **21**, 51–78.
- Jury, M. R., 1996: Regional teleconnection patterns associated with summer rainfall over South Africa, Namibia and Botswana. *Int. J. Climatol.*, **16**, 135–153.
- , H. M. Mulenga, and S. J. Mason, 1999: Exploratory long-range models to estimate summer climate variability over southern Africa. *J. Climate*, **12**, 1892–1899.
- Karl, T. R., W.-C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman, 1990: A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *J. Climate*, **3**, 1053–1079.
- Kumar, A., and M. P. Hoerling, 2000: Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 255–264.
- Landman, W. A., and S. J. Mason, 1999a: Change in the association between Indian Ocean sea-surface temperatures and summer rainfall over South Africa and Namibia. *Int. J. Climatol.*, **19**, 1477–1492.
- , and —, 1999b: Operational long-lead prediction of South African rainfall using canonical correlation analysis. *Int. J. Climatol.*, **19**, 1073–1090.
- , and W. J. Tennant, 2000: Statistical downscaling of monthly forecasts. *Int. J. Climatol.*, **20**, 1521–1532.
- , S. J. Mason, P. D. Tyson, and W. J. Tennant, 2001a: Retroactive skill of multi-tiered forecasts of summer rainfall over southern Africa. *Int. J. Climatol.*, **21**, 1–19.
- , —, and —, 2001b: Statistical downscaling of GCM simulations to streamflow. *J. Hydrol.*, **252**, 221–236.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- Makarau, A., and M. R. Jury, 1997: Predictability of Zimbabwe summer rainfall. *Int. J. Climatol.*, **17**, 1421–1432.
- Mason, S. J., 1998: Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *Int. J. Climatol.*, **18**, 147–164.
- , and A. M. Joubert, 1997: Simulated changes in extreme rainfall over southern Africa. *Int. J. Climatol.*, **17**, 291–301.
- , and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725.
- , A. M. Joubert, C. Cosijn, and S. J. Crimp, 1996: Review of seasonal forecasting techniques and their applicability of southern Africa. *Water SA*, **22**, 203–209.
- , L. Goddard, N. E. Graham, E. Yelaeva, L. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853–1873.
- Mattes, M., and S. J. Mason, 1998: Evaluation of seasonal forecasting for Namibian rainfall. *S. Afr. J. Sci.*, **94**, 183–185.
- Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Appl. Meteor.*, **26**, 1589–1600.
- Mjelde, J. W., T. N. Thompson, C. J. Nixon, and P. J. Lamb, 1997: Utilising a farm-level decision model to help prioritise future climate prediction research needs. *Meteor. Appl.*, **4**, 161–170.
- Murphy, A. H., 1998: The early history of probability forecasts: Some extensions and clarification. *Wea. Forecasting*, **13**, 5–15.
- Palmer, T. N., and D. L. T. Anderson, 1994: The prospects of seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.*, **120**, 755–793.
- Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, **282**, 728–731.
- Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.
- Solman, S. A., and M. N. Nuñez, 1999: Local estimates of global climate change: A statistical downscaling approach. *Int. J. Climatol.*, **19**, 835–861.
- Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373.
- Trenberth, K. E., and J. W. Hurrell, 1994: Decadal atmosphere–ocean variations in the Pacific. *Climate Dyn.*, **9**, 305–319.
- Tyson, P. D., and R. A. Preston-Whyte, 2000: *The Weather and Climate of Southern Africa*. Oxford, 396 pp.
- Ward, M. N., and C. K. Folland, 1991: Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. *Int. J. Climatol.*, **11**, 711–743.
- Wilby, R. L., and T. M. L. Wigley, 2000: Precipitation predictors for downscaling: Observed and general circulation model relationships. *Int. J. Climatol.*, **20**, 641–661.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489.